Set of facts closed under a schema and update operations associated

Anne Verroust

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(3)954 90 20

Rapports de Recherche

N° 285

SET OF FACTS CLOSED UNDER A SCHEMA AND UPDATE OPERATIONS ASSOCIATED

Anne VERROUST

Avril 1984

SET OF FACTS CLOSED UNDER A SCHEMA

AND UPDATE OPERATIONS ASSOCIATED

by Anne VERROUST

Abstract :

An object is a set of attributes corresponding to a real life sentence. We propose a model where the elementary elements of information are defined on objects. We call them facts.
From a set of facts we perform deduction of new facts using two deduction rules : one based on projection and the other using join. The "compatibility relation" between objects defines the possible joins.
We then study update operations in such a context. This led us to characterise the "good schemata" for the insertion and the deletion. This model is also interpreted as a representation of an incomplete relation. The constraints induced by the deduction rules are then compared with the classical notion of decomposability.

Resumé :

Un objet est un ensemble d'attributs correspondant à une phrase de la vie réelle. Nous proposons un modèle où les éléments d'information sont définis sur les objets. Ces éléments sont appelés faits.
Nous introduisons deux règles permettant de déduire de nouveaux faits d'un ensemble de faits : l'une utilisant la projection et l'autre la jointure. Les jointures permises sont déterminées par la "relation de compatibilité" définie sur les objets.
Nous étudions les opérations de mise à jour dans un tel contexte. Cette étude conduit à une caractérisation des "bons schémas" pour les opérations d'insertion et de suppression. Ce modèle est également interprété en tant que représentation d'une relation incomplète. Les contraintes induites par les règles de déduction sont alors comparées avec la notion classique de décomposabilité.

INTRODUCTION

Consider the following relation :

R :	CLASS	COURSE	STUDENT	TEACHER
	6	English	John	Smith
	5	English	Albert	Smith

and assume we want to insert :

"Jane takes the English course in the class 6" without knowing the teacher,

or we want to delete :

"Smith teaches English" without loosing the information relative to his students.

We then are faced with two different problems :

(i) the definition of the result of such updates
(ii) the computation of this result.

The classical avoid update anomalies is to decompose the relation and compute the updates in the decomposed parts of the relation.

In this particular example R would be decomposed in :

R_1 = R[CLASS,COURSE,STUDENT], R_2 = R[CLASS,COURSE,TEACHER],
R_3 = R[CLASS,STUDENT], R_4 = R[CLASS,TEACHER],
R_5 = R[COURSE,STUDENT], R_6 = R[COURSE,TEACHER]

(i.e. $R = R_1 * \dots * R_6$).

If we delete "Smith teaches English" from R_6 then we delete all the information relative to his students.

Thus, extending this solution, we could consider that the avalaible information is contained in all the possible join paths built from $R_1 \ldots R_6$ i.e :

$R_1 * R_2$, $R_1 * R_3$, $R_1 * R_4$, $R_1 * R_5$, $R_1 * R_6$, $R_2 * R_3$,
$R_2 * R_4$, $R_2 * R_5$, $R_2 * R_6$, $R_3 * R_4$, $R_5 * R_6$, etc...

(As in Biskup and Bruggeman [BiB], we do not consider cartesian products).

In the case of the above example we know that :

"John learns English"

and

"Smith teaches English in the class 5"

We conclude that

"John follows the English course of Smith in the class 5"

which may not be true.

Therefore we must determine which are the possible joins : it is the role of the "compatibility relation".

In this case, we will just take into account :

R_1, R_2, R_3, R_4, R_5, $R_1 * R_2$, $R_1 * R_3$, $R_1 * R_5$, $R_2 * R_4$, $R_2 * R_6$,
$R_1 * R_3 * R_2$, $R_1 * R_5 * R_2$, $R_1 * R_2 * R_4$, $R_1 * R_2 * R_6$

(A teacher teaches to all the students of a given course in a given class).

A query on CLASS STUDENT will be evaluated on the union of R_1, $R_1 * R_2$, R_3, $R_1 * R_3$, $R_1 * R_5$, $R_1 * R_3 * R_2$, and $R_1 * R_5 * R_2$.

To be consistent, we want to get the same result while evaluating the request on R_3 = R[CLASS,STUDENT].

This leads to the notion of a closed set of facts.

Then we shall investigate the problem of defining and computing updates while keeping this closure constraint. This yields a characterization of well behaved schemata.

The model

Let us consider a set of attributes U. Some subsets of U have a particular property : they correspond to predicates from real life. Let us call them objects similarly to Sciore's definition of objects [Sci]. The set of all the objects included in U is a covering of U. Let us denote it $\underline{0}$.

By definition, any sentence using attributes of U will use all the attributes of an object. Thus any element of information concerning some attributes of U will correspond to a tuple defined on some object. This leads to the notion of fact : Given the set of objects $\underline{0}$, a fact is a tuple defined on an element of $\underline{0}$ (i.e. an application $x : X \rightarrow \bigcup_{A \in X} D(A)$ where $X \in \underline{0}$ s.t x(A) belongs to the domain associated with A i.e D(A) for any A in X).

Example 1.

Let U be U = {CLASS,COURSE,STUDENT,TEACHER,ROOM} that we shall abbreviate U = {Cl,Co,S,T,R}. The set of associated objects is :

- CLASS COURSE STUDENT TEACHER ROOM
- CLASS COURSE STUDENT TEACHER
- CLASS COURSE STUDENT ROOM
- CLASS COURSE STUDENT
- CLASS COURSE TEACHER ROOM
- CLASS COURSE TEACHER
- CLASS STUDENT

- CLASS TEACHER
- COURSE STUDENT
- COURSE TEACHER

For example,

- CLASS COURSE STUDENT ROOM corresponds to the sentence "Student S of class Cl takes the course Co in the ROOM R"

- CLASS COURSE TEACHER corresponds to "Teacher T teaches course Co in the class Cl".

But TEACHER STUDENT is not an object. The sentence corresponding to this set of attributes is :

"There exist a class Cl and a course Co where student S is a student and teacher T is a teacher"

It uses two more attributes CLASS and COURSE. Thus TEACHER STUDENT does not correspond to predicate from real life.

Given two facts, we can sometimes deduce a new information by joining them :
Let us take the set of attributes U of the example 1. If we know that :

Albert follows the English courses in the class 5

and

Smith teaches the English courses in the class 5

then we can deduce that :

Smith teaches English to Albert in the class 5

(We assume then that all the students of a given course in a given class have the same teachers).

But sometimes we cannot deduce any information by joining them :
If we only know just that :

"Albert follows the English courses"

and

"Smith teaches English"

We cannot ensure that

"Smith teaches English to Albert"

(We have to know the classes where Smith is a teacher and where Albert is a student to ensure that).

Thus we need to know which are the possible join between facts. The <u>compatibility relation</u> @ between objects gives us this information : if two objects are compatible then we will be able to deduce a new fact by joining two facts defined on these objects.

This compatibility relation @ must satisfy :

(1) if X @ Y then $X \cap Y \neq \emptyset$ and $X \cup Y \in \underline{O}$
because the result of a join must correspond to a real life sentence

(2) If X, Y are two objects such that $X \subseteq Y$ then X @ Y

(3) @ is symmetric : if X @ Y then Y @ X

Note that by condition (1) the set of objects $\underline{O}$ is closed under union of compatible objects.

Example 2 : In the previous example we have :

$\underline{O}$ = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR,CLCoST,ClCoSTR}

The compatibility relation @ associated to <u>0</u> is defined as follows :

@ satisfies condition (2) and (3) and

. (ClCoS) @ (ClCoT)

. (ClCoSR) @ (ClCoTR),(ClCoS) @ (ClCoTR),(ClCoSR) @ (ClCoT)

Intuitively, this compatibility relation describes the embedded multivalued dependencies

$$\text{if } X \; @ \; Y \text{ then } X \cap Y \twoheadrightarrow (X-Y)/(Y-X) \text{ on } X \cup Y$$

Knowing this compatibility relation, we can now define the possible joins between two facts :

<u>Definition 1 :</u>

Let X and Y be two compatible objects w.r.t. @. Then two facts x and y over X and Y such that $x_{|X \cap Y} = y_{|X \cap Y}$ are said to be <u>compatible</u> and in this case $x * y$ is the fact over $X \cup Y$ defined by

$$(x * y)_{|X} = x \text{ and } (x * y)_{|Y} = y \; .$$

This join operation is extended to sets of facts as follows : If **A** and **B** are two sets of facts then

$$\mathbf{A} * \mathbf{B} = \{x * y \mid (x \in \mathbf{A}) \wedge (y \in \mathbf{B}) \wedge (x \text{ and } y \text{ compatible})\}$$

Given the compatibility relation @, we can determine the set of all the facts deduced by join from a finite set of fact **A**. We call it $\mathbf{A}^*$:

Let **A** be a finite set of facts, $\mathbf{A}^*$ is the limit of the sequence $\mathbf{A}_n$ defined by $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_{n+1} = \mathbf{A}_n * \mathbf{A}$. (This sequence converges after a finite number of steps since it is non decreasing and has an upperbound).

<u>**Note that :**</u> . $\mathbf{A}^*$ is closed under join operation

. All the facts of $\mathbf{A}^*$ are defined on $\underline{0}$.

Some facts contain more information than other facts (in the same sense of Zaniolo [Za]). This leads to the following order between facts :

<u>**Definition 2 :**</u>

Let x and y be two facts over X and Y

$$x \geqq y \text{ iff } X \supseteq Y \text{ and } x_{|Y} = y$$

These can be some redundancy in the information contained in $\mathbf{A}^*$. $\bar{\mathbf{A}}$ is the set of facts which contains the same information than $\mathbf{A}^*$ but which is not redundant w.r.t. $\leqq$:

$\bar{\mathbf{A}}$ is the set of the maximal elements for $\geqq$ of $\mathbf{A}^*$

$$\bar{\mathbf{A}} = \{x \in \mathbf{A}^* \mid \forall y \in \mathbf{A}^*, (y \geqq x) \Rightarrow (y = x)\}$$

moreover

$\mathbf{A} \subseteq \mathbf{A}^*$ and $\bar{\mathbf{A}} \subseteq \mathbf{A}^*$,

$\forall x \in \mathbf{A}, \exists\, y \in \bar{\mathbf{A}}$ s.t. $y \geqq x$ but in general $\mathbf{A} \not\subseteq \bar{\mathbf{A}}$

As we can deduce now facts using join, we can restrict the set of objects where the facts are defined to avoid redundancy of information. First we have to introduce the notion of @-connectivity :

<u>**Definition 3 :**</u>

$C \subseteq \underline{0}$ is <u>@-connected</u> if there exist a numbering $X_1 \ldots X_n$ of C such that $\forall\, i \geqq 2$, $X_i \mathbin{@} \bigcup_{j<i} X_j$.

(if @ satisfies : $\forall\, X, Y \in \underline{0}$, $X \mathbin{@} Y \Leftrightarrow X \quad Y \neq \emptyset$ we obtain the notion of connectivity of the hypergraphs [Ber]).

Thus if C is @ connected, there exist a numbering of C, $X_1 \dots X_n$ and a sequence of facts $x_1, \dots x_n$, each x_i defined on X_i, such that $(\dots(x_1 * x_2) * \dots * x_n)$ is a fact on $\bigcup_{i \leq n} X_i$.

We can restrict our facts to be defined on the minimal subset of $\underline{0}$ (if it exists) which can generate all the objects.

Definition 4 :

Given $\underline{0}$ and @, the schema S is the minimal subset of $\underline{0}$ such that any element X of $\underline{0}$ is covered by an @-connected subset C of S : $X = \bigcup_{Y \in C} Y$

The elements of S will be called atoms and the facts defined on these atoms atomic facts.

Example 3 :

Let us take the set of objects and the compatibility relation of the example 2.

$\underline{0}$ = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR,ClCoST,ClCoSTR}

Then S = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR}

- ClCoST is the union of the two compatible objects ClCoS and ClCoT

  and ClCoSTR is the union of ClCoTR and ClCoS

- Thus any element of $\underline{0}$ is covered by an @-connected subset of S.

S is minimal because any element X of S cannot be covered by a @-connected subset of S-{X}.

Let us note thet :

- S is uniquely determined by the set of objects $\underline{0}$ and the compatibility relation @ : It is composed of the objects X such that if Y @ Z and X = Y $\cup$ Z then X=Y or X=Z (X cannot be decomposed in two other compatible objects).

- $\underline{0}$ is uniquely determined by the schema S and the compatibility relation @ :

  X $\in$ $\underline{0}$ iff there exist a subset C of S @-connected such that $X = \bigcup_{Y \in C} Y$

- But if we know only S and $\underline{0}$ we cannot know in general the compatibility relation @ :

  If S = {AB,BC,AC}
  and $\underline{0}$ = {AB,BC,AC,ABC}

  We have many different compatibility relations having S as schema and $\underline{0}$ as set of objects :

  $@_1$: AB $@_1$ BC, AB $@_1$ ABC, AC $@_1$ ABC, BC $@_1$ ABC

  $@_2$: AB $@_2$ AC, AB $@_2$ ABC, AC $@_2$ ABC, BC $@_2$ ABC

  $@_3$: BC $@_3$ AC, AB $@_3$ ABC, AC $@_3$ ABC, BC $@_3$ ABC

  for example.

We now define the projection of a set of facts on a subset of $\underline{0}$:

Definition 5 :

Let **A** be a set of facts and W a subset of $\underline{0}$. The projection of **A** on W is :

$$\pi_W(\mathbf{A}) = \{x \mid \exists y \in \mathbf{A}, \exists X \in W \text{ s.t. } x = y_{|X}\}$$

If **A** is a finite set of atomic facts and S the schema :

- $\pi_S(\mathbf{A}^*) = \pi_S(\bar{\mathbf{A}})$

- $\pi_S(\mathbf{A}^*) \supseteq \pi_S(\mathbf{A}) \supseteq \mathbf{A}$

The compatibility relation between objects and the order relation $\leq$ between facts induce two deduction rules for facts :

(a) Deduction by join :

If x and y are two compatible facts then the knowledge of x and y implies the knowledge of $x * y$

(b) Deduction by projection :

If x and y are facts such that $x \leq y$ then the knowledge of y implies the knowledge of x.

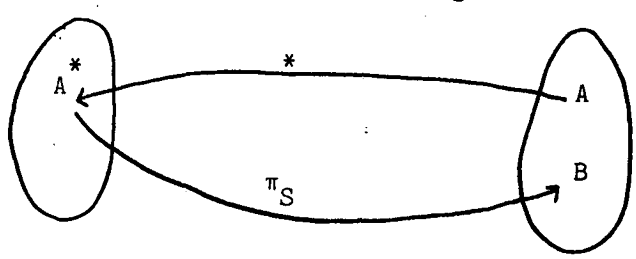
Thus, as we allow these two deduction rules, the information contained in a set of facts **A** is the closure of this set under join and projection on S.

We want the information to be totally explicit on the atoms (i.e. no new atomic fact can be deduced using join and projection). Thus we introduce the notion of "closure under S" :

Definition 6 :

A is closed under S (given @) iff $\pi_S(\mathbf{A}^*) = \mathbf{A}$

sets of facts



sets of atomic facts

A is closed under S iff **A** = **B**

Then if **A** is closed under S we cannot deduce new atomic facts from **A** using join and projection.

Example 4 :

Let us take U, 0, @ and S as in the example 3.

A :

CLASS	COURSE	STUDENT	TEACHER	ROOM
6	English		Smith	
6	English	Paul		
6	English	Paul		33
6		Jane		

$\mathbf{A}^*$:

CLASS	COURSE	STUDENT	TEACHER	ROOM
6	English		Smith	
6	English	Paul		
6	English	Paul		33
6		Jane		
6	English	Paul	Smith	
6	English	Paul	Smith	33

and $\bar{\mathbf{A}}$:

CLASS	COURSE	STUDENT	TEACHER	ROOM
6	English	Paul	Smith	33
6		Jane		

In this example $\mathbf{A} \not\subseteq \bar{\mathbf{A}}$

$\pi_S(\mathbf{A}^*) = \pi_S(\bar{\mathbf{A}})$:

CLASS	COURSE	STUDENT	TEACHER	ROOM
6		Paul		
6		Jane		
6			Smith	
	English	Paul		
	English		Smith	
6	English	Paul		
6	English		Smith	
6	English	Paul		33
6	English		Smith	33

In this example $\mathbf{A} \neq \pi_S(\mathbf{A}^*)$.

Let us summarize the notions introduced in this section :

A set of objects $\underline{O}$ and a compatibility relation @ associated define a schema S. We allow two deduction rules : one based on join and the other one on projection under the schema S. We restrict our interest to sets of atomic facts closed under join and projection on S.

In the next section we study the problem of defining and computing the result of update operations which keep the property of "closure under S".

The update problems

1-Inserting an atomic fact

Let **A** be a set of facts closed under S and x an atomic fact. We want to preserve the closure property while inserting x in **A**. In general $\mathbf{A} \cup \{x\}$ is not closed under S :

Example 5 :

Let U be {EMPLOYEE,DEPARTMENT,MANAGER} that we shall abbreviate {E,D,M}

$\underline{O}$ is {EM,ED,MD,EMD}

We suppose that each manager of a department is the manager of all the employees.

Then @ is defined by

$$\forall X, Y \in \underline{O} \quad X \mathbin{@} Y \Leftrightarrow X \cap Y \neq \emptyset$$

Thus S is {EM,ED,MD}

Let **A** be

EMPLOYEE	MANAGER	DEPARTMENT
Albert	Shoes	
Paul	Shoes	
Jane	Books	
Jane		Max
	Books	Max

Suppose we want to insert "Brown is Paul's manager" $\mathbf{B} = \mathbf{A} \cup \{\langle \text{Paul Brown} \rangle\}$ is not closed under S because $\pi_S(\mathbf{B}^*) = \mathbf{B} \cup \{\langle \text{Shoes Brown} \rangle\} \neq \mathbf{B}$.

Thus by adding a new fact x in **A** we may add implicitely (allowing deduction by join and projection on S) some new information that is not included in $\mathbf{A} \cup \{x\}$.

We have to determine the side effects produced by the addition of x in **A.** The new facts implicitely added, during this insertion are included in all the sets of facts **B** closed under S and containing $\mathbf{A} \cup \{x\}$.

A reasonnable definition of the result of the insertion of x in **A** should be : the intersection of all the sets of facts **B** closed under

S and containing $\mathbf{A} \cup \{x\}$ or equivalently, the minimal set of facts $\mathbf{B}$ closed under S such that $\mathbf{A} \cup \{x\}$ is included in $\mathbf{B}$. This minimum exists, and is computable :

Theorem 1 :

For any set of atomic facts $\mathbf{A}$, there exist a unique minimal set of atomic facts $\mathbf{A}^S$ such that

(1) $\mathbf{A} \subseteq \mathbf{A}^S$

(2) $\mathbf{A}^S$ is closed under S

The proof follows directly from the theorem :

Theorem 2 :

For any set of atomic facts $\mathbf{A}$, $\mathbf{A}^S$ is the limit of the sequence $(\mathbf{A}^n)_{n\in\mathbf{N}}$ defined by

$$\mathbf{A}^o = \mathbf{A} \text{ and } \mathbf{A}^{n+1} = \pi_S\ ((\mathbf{A}^n)^*)$$

Proof :

(a) We first show that $(\mathbf{A}^n)_{n\in\mathbf{N}}$ is stationnary.

. For a given $\mathbf{B}$, $\mathbf{B} \subseteq \mathbf{B}^*$ and π_S is non decreasing thus $(\mathbf{A}^n)_{n\in\mathbf{N}}$ is an increasing sequence.

. For any integer n, and any attribute A in U the values of A appearing in the facts of $\mathbf{A}^n$ are in $V(A) = \{x(A) \mid (x(\mathbf{A}) \wedge (x \text{ is defined on } A)\}$.
$\mathbf{A}$ is finite, thus V(A) is finite.
Let $\mathbf{B}$ be the set of facts defined as follows :
For each X in S, $X = \{A_1 \ldots A_k\}$ with $V(A_i) \neq \emptyset\ \forall\ i$,
For each $(a_1 \ldots a_k)$ in $V(A_1) \times \ldots \times V(A_k)$

$$x : X \to \bigcup_{A \in X} D(A) \qquad \text{is in } \mathbf{B}$$
$$A_i \to a_i$$

B is a finite set of facts such that $\mathbf{A}^n \subseteq \mathbf{B}$ for any integer n.

Thus there exist N such that $\mathbf{A}^N = \mathbf{A}^{N+k}$, $\forall\, k \in \mathbf{N}$

for k=1, $\mathbf{A}^N = \mathbf{A}^{N+1} = \pi_S(\mathbf{A}^{N*})$

Thus $\mathbf{A}^N$ is closed under S.

(b) Let **B** be a set of facts closed under S such that $\mathbf{B} \supseteq \mathbf{A}$.

We have $\mathbf{B} = \pi_S(\mathbf{B}^*)$,(**B** is closed under S)

$\supseteq \pi_S(\mathbf{A}^*)$ $(\mathbf{B} \supseteq \mathbf{A})$

Then, by induction on n we have

$\mathbf{B} \supseteq \mathbf{A}^n \quad \forall\, n \in \mathbf{N}$

Thus $\mathbf{B} \supseteq \mathbf{A}^N$ and $\mathbf{A}^N = \mathbf{A}^S$ Q.E.D.

Going back to example 4 we have :

$\mathbf{B} = \mathbf{A} \cup$ {<Paul Brown>}

$\mathbf{B}^1 = \pi_S(\mathbf{B}^*) = \mathbf{B} \cup$ {<Shoes Brown>}

$\mathbf{B}^2 = \pi_S(\mathbf{B}^{1*}) = \mathbf{B}^1 \cup$ {<Albert Brown>}

$\mathbf{B}^3 = \pi_S(\mathbf{B}^{2*}) = \mathbf{B}^2$

i.e. $\mathbf{B}^S = \mathbf{B}^2$:

EMPLOYEE	DEPARTMENT	MANAGER
Albert	Shoes	
Paul	Shoes	
Jane	Books	
Paul		Brown
Albert		Brown
Jane		Max
	Shoes	Brown
	Books	Max

In this example we obtain the result of the insertion in two steps.

The result of the insertion is computed using the sequence $\mathbf{B}^n$ of the previous theorem. This sequence reaches its limit in a finite number of steps but this number is not bounded in general. The following theorem prove this assertion :

Theorem 3 :

- If U contains the attributes having an infinite domain of values

or

- If U is infinite. Then

For any integer n, there exist a schema S_n a compatibility relation $@_n$ and an associated set of atomic facts **A** such that

$$\mathbf{A}^{S_n} \neq \mathbf{A}^n$$

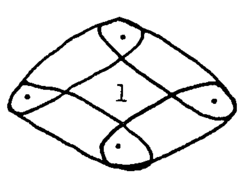
Proof :

Let the compatibility relation @ be :

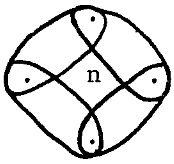
$X, Y \in \underline{O} \quad X @ Y \iff X \cap Y \neq \emptyset$

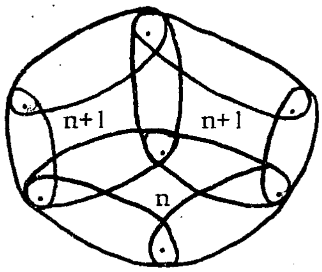
(A) Suppose U infinite

$\forall\, n \in \mathbf{N}$, we construct S_n recursively as follows :

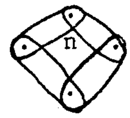
S_1 is 

S_{n+1} is obtained from S_n by transforming each

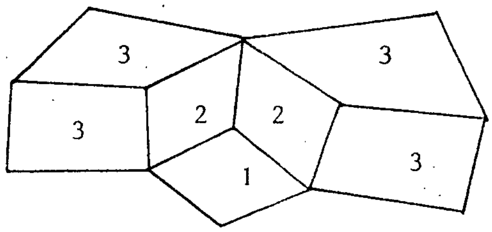




in adding 3 new vertices and 5 new edges to each



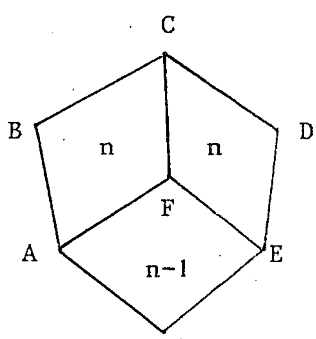
for example



we represent it by a graph since all the same cardinality = 2.

Given S_n, **A** set of atomic facts such that $\mathbf{A}_n$ is not closed under S_n is defined as follows :

- for each couple n-n (when n > 1) :



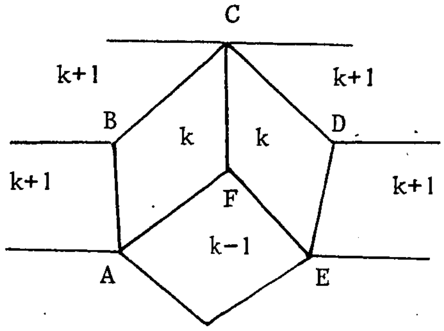
$t_n : \left| \begin{array}{l} B \to 1 \\ A \to 1 \end{array} \right. \qquad t'_n : \left| \begin{array}{l} C \to 2 \\ F \to 1 \end{array} \right.$

$r_n : \left| \begin{array}{l} B \to 1 \\ C \to 1 \end{array} \right. \qquad r'_n : \left| \begin{array}{l} C \to 2 \\ D \to 2 \end{array} \right.$

$s_n : \left| \begin{array}{l} C \to 1 \\ F \to 1 \end{array} \right. \qquad s'_n : \left| \begin{array}{l} D \to 2 \\ E \to 2 \end{array} \right.$

are in **A**.

. For each couple k-k 1 < k < n :

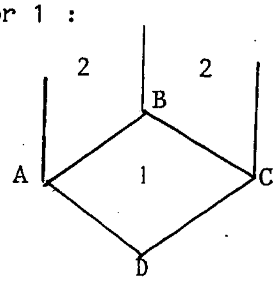


t_k : | C → 2
| F → 1

r_k : | C → 1
| F → 1

are in **A**

. For 1 :



t_1 : | C → 2
| D → 1

$t_1 \in \mathbf{A}$

Then t | A → 1 $\quad t \in \mathbf{A}_{n+1}$
| D → 1 $\quad t \notin \mathbf{A}_n$ □

(B) Suppose that there exist three attributes A, B, C in U such that D(A), D(B) and D(C) are infinite.
We will show that S = {AB,BC,AC} does not converge in n steps for any integer n.

(We assume, for simplicity that D(A), D(B) and D(C) contain **N**). Let n be an integer, n > 1. Then the set of atomic facts **A** defined as follows is such that $\mathbf{A}_n$ is not closed under S :

on AB | we have $\{(i,j) \mid |i-j| \leq 1$ and $\{i,j\} \subset \{1,\ldots,2^{n+1}+1\}\}$
BC |
AC |

Then $\pi_S(\mathbf{A}^*) = \mathbf{A}_1$ is composed of the facts

$$\left.\begin{array}{l} \text{on} \quad AB \\ \quad\quad\ BC \\ \text{and} \quad AC \end{array}\right| \quad \{(i,j) \mid |i-j| \leq 2 \text{ and } \{i,j\} \subset \{1 \ldots 2^{n+1}+1\}\}$$

and $\mathbf{A}_k$ is composed of the facts, for any $k \leq n+1$

$$\left.\begin{array}{l} \text{on} \quad AB \\ \quad\quad\ BC \\ \quad\quad\ AC \end{array}\right| \quad \{(i,j) \mid |i-j| \leq 2^k \text{ and } \{i,j\} \subset \{1 \ldots 2^{n+1}+1\}\}$$

Thus $\mathbf{A}_n \neq \mathbf{A}_{n+1}$ and $\mathbf{A}_n$ is not closed under S. Q.E.D.

From a computational point of view it is important to characterize the cases where the sequence converges in few steps. (Note that the number of steps conditions the number of joins and projection necessary to compute the result).

We shall say that the schema S <u>converges in n steps</u> iff for any set of atomic facts $\mathbf{A}$, $\mathbf{A}^S$ is computed at most in n steps (i.e. $\mathbf{A}^S = \mathbf{A}^n$ where $\mathbf{A}^0 = \mathbf{A}$ and $\mathbf{A}^{n+1} = \pi_S((\mathbf{A}^n)^*)$. We will give a characterisation of this property for n=0 or 1.

In order to give these characterisations we introduce the notion of <u>minimum precycle description</u> of an atom :

<u>Definition 7 :</u>

Given S and @, if X is an atom <u>a minimum precycle description</u> of X is a subset C of S-{X} such that :

(i) C is @-connected

(ii) $X \subseteq \bigcup_{Y \in C} Y$

(iii) C is minimum in the following way : every proper subset of C fails to satisfy ((i) and (ii))

MPC(X) will be the (possibly empty) set of the minimum precycle description of X.

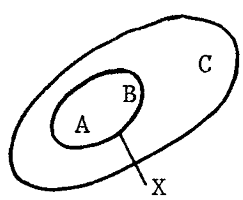
Example 6 :

(i) Let us take the schema S and the compatibility relation @ of the example 3 :

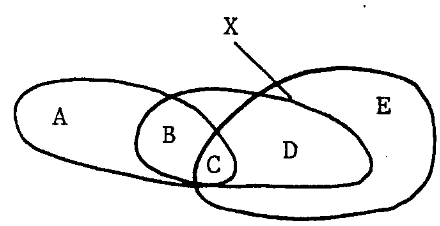
S = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR}

MPC(ClCoSR) = {{ClCoS,ClCoTR}}
MPC(ClS) = {{ClCoS},{ClCoSR}}
MPC(ClCoS) = {{ClCoSR}}

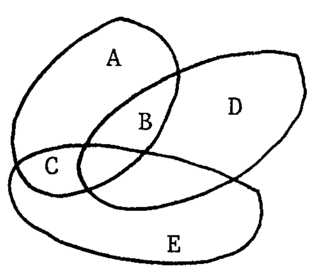
(ii) If @ satisfies X @ Y <=> $X \cap Y \neq \emptyset$ then S can be represented by a hypergraph.
In this case,



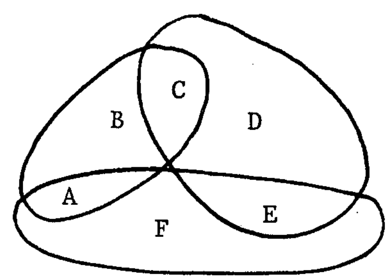
MPC(X) = {ABC}



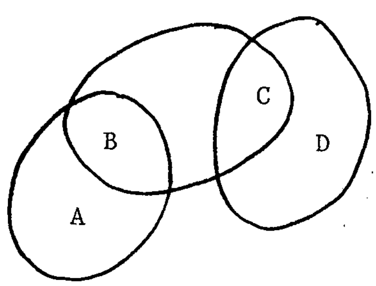
MPC(X) = {{ABC,CDE}}



MPC(X) = $\emptyset$ $\forall$ X $\in$ S



MPC(X) = $\emptyset$ $\forall$ X $\in$ S



MPC(X) = $\emptyset$ $\forall$ X $\in$ S

Let us now characterize the schemas S which converge in 0 step.

Lemma : Let **A** be a set of atomic facts and x an atomic fact defined on X. If **A** contains a finite sequence $x_1 \dots x_n$ of facts such that $x \leq (\dots(x_1 * x_2)\dots * x_n)$ and $x \neq x_i \ \forall i$ then $MPC(X) \neq \emptyset$.

Proof : Let us take a minimal sequence $y_1 \dots y_k$ in **A** such that $x \leq (\dots(y_1 * y_2)\dots * y_k)$ and $y_i \neq x \ \forall i$.
If each y_i is defined on Y_i then
$\{Y_1 \dots Y_k\}$ is @ connected (because $(\dots(y_1 * y_2)\dots * y_k)$ exists)
and $X \subseteq \bigcup_{i \leq k} Y_i$ (by the definition of $\leq$).

Then, by the minimality of $y_1 \dots y_k$, $\{Y_1 \dots Y_k\} \in MPC(X)$.
Q.E.D.

Theorem 4 :

S converges in 0 step iff

$$\forall X \in S, \quad MPC(X) = \emptyset$$

Proof :

(a) Suppose that S does not converge in 0 step. Then there exist a finite set of atomic facts **A** such that $\mathbf{A} \neq \pi_S(\mathbf{A}^*)$.
Let $x \in \pi_S(\mathbf{A}^*) - \mathbf{A}$. x is defined on X.
There exist a sequence $x_1 \dots x_n$ of atomic facts included in **A** such that

$$x \leq (\dots(x_1 * x_2)\dots * x_n) \text{ and } x_i \neq x \quad \forall i$$

then by the lemma, $MPC(X) \neq \emptyset$.

(b) Conversely, assume that :

$\exists\ X \in S,\ \exists\ C \subset S,\ C \in MPC(X).$

Let u be a fact defined on $\bigcup_{Y \in C} Y$ and

$\mathbf{A} = \bigcup_{Y \in C} \{u|_Y\}$

then $\mathbf{A}^*$ contains u (C is connected)
and $\pi_S(\mathbf{A}^*)$ contains $u|_X$ which is not in $\mathbf{A}$.
Thus S does not converge in 0 step. Q.E.D.

By knowing the minimum precycle description of X for all the X in S, we can restrict the number of joins when computing $\pi_S(\mathbf{A}^*)$:

Theorem 5 :

If $n = \sup \{ |C| \ | \exists\ X \in S,\ C \in MPC(X)\}$ then for any set of atomic facts $\mathbf{A}$,

$$\pi_S(\mathbf{A}^*) = \pi_S(\mathbf{A}_n) \quad \text{where } \mathbf{A} \text{ belongs to the sequence } (\mathbf{A}_k)_{k \in \mathbf{N}}$$

defined by $\mathbf{A}_1 = \mathbf{A}$ and $\mathbf{A}_{k+1} = \mathbf{A}_k * \mathbf{A}$.

Proof :

- $\mathbf{A}_k$ is an increasing sequence and π_S is non decreasing thus $\pi_S(\mathbf{A}^*) \supseteq \pi_S(\mathbf{A}_k)$ for any set of atomic facts $\mathbf{A}$.

- For any x in $\pi_S(\mathbf{A}^*)$ lets take a minimum sequence $x_1 \ldots x_k$ of $\mathbf{A}$ such that $x \leq (\ldots(x_1 * x_2)\ldots * x_k)$ each x_i beeing defined on X_i.
  Then $\{X_1 \ldots X_k\} \in MPC(X)$ and $k \leq n$
  Thus $(\ldots(x_1 * x_2)\ldots * x_k) \in \mathbf{A}_n$. Q.E.D.

We now characterise the schema S that converge in 1 step :

Theorem 6 :

The schema S converges in 1 step iff (I) is not satisfied :

(I) There exist X in S, C_X in MPC(X) and $C \subset C_X$, $C \neq \emptyset$ and for each Y in C, C_Y in MP(Y) such that

If $C' \in PCM(X)$ is such that

$$C' \subseteq C_X \cup (\bigcup_{Y \in C} C_Y)$$

$$\text{and } C' \cap (\bigcup_{Y \in C} C_Y) \neq \emptyset$$

then,

either $\exists\ Y \in C,\ Z \in (C_X - \{Y\}) \cup \{X\}$ such that

$$Z \cap Y \neq Z \cap (\bigcup_{V \in C' \cap C_Y} V)$$

or $|C| \geq 2$ and $\exists$ Y, Z in C $Y \neq Z$ such that

$$Z \cap Y \neq (\bigcup_{V \in C' \cap C_Y} V) \cap (\bigcup_{V \in C' \cap C_Z} V)$$

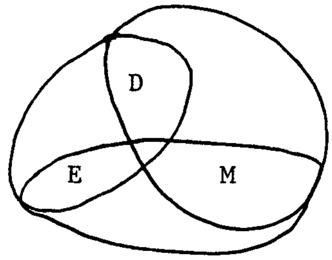
The proof is given in appendix.

Intuitively when (I) is satisfied there are cases where a new fact defined on X is obtained after two steps at least

- After the first one we get the facts defined on the $Y \in C$. These facts will be used in the second step to obtain a new fact on X
  The condition on the Y, C' and eventually on Z makes it impossible to obtain this new fact on X in only one step (one step means one computation of closure under join followed by projection on S).

Example 7 :

a) The schema S with the compatibility relation of the example 5 does not converge in 1 step :



Lets take $X = \{ED\}$

$MPC(X) = \{\{DM, ME\}\} = \{C_X\}$

$C = \{DM\}$

$MPC(DM) = \{\{ED, EM\}\} = \{C_Y\}$

there is only one C' such that

$$C' \subset C_X \cup C_Y,\ C' \in MPC(X) \text{ and } C' \cap C_Y \neq \emptyset$$

it is $C' = C_X$

and we have $Y = DM \in C$

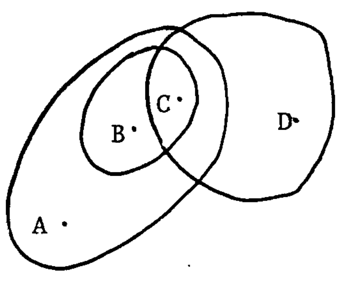
$Z = ME \in C_X - \{Y\}$

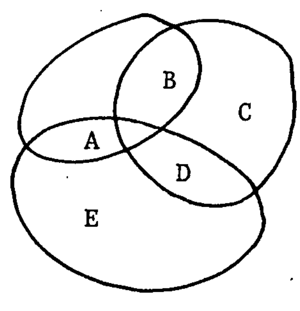
$Z \cap Y = \{M\}$ and $Z \cap (ED \cup EM) = DM$

Thus condition (I) is satisfied.

b) When @ satisfies $X @ Y \Leftrightarrow X \cap Y \neq \emptyset$
S can be represented by a hypergraph.

(I) is not satisfied





Corollary :

If S satisfies

a) $\forall X \in S,\ C \in MPC(X) \Rightarrow |C| = 1$
then S converges in one step.

Proof :

If S satisfies the condition (I) of the previous theorem then there exist X in S where C_X has two elements ($C \subset C_X$ is non trivial).
Q.E.D.
Thus (a) cannot be satisfied.

In this section we have defined an insertion operation which preserves the closure property under a schema. We have shown that the result of an insertion is uniquely defined. We have shown that, in the general case, the number of joins and projections necessary to compute the result is unbounded and we have characterized the sets of atoms where this number is limited to 0 and 1.

These two cases are the most interesting ones :

- When S converges in 0 step, the definition of the insertion is the traditional one.

- When S converges in one step, to compute the result of an insertion, we have to compute the closure under join and project the result on S. This computation is exactly what we have to do to verify if a set of atomic facts is closed under S.

Deletion of an atomic fact

Consider a set of facts **A** closed under S and a fact x belonging to it. We want to delete from **A** the information related to x while keeping the property of closure under S. We run into the same problem than in the case of an insertion : in general **A**-{x} is not closed under S :

Example 8 :

Let us take the same set of atomic facts **A** than in example 5 :

A :

EMPLOYEE	MANAGER	DEPARTMENT
Albert	Shoes	
Paul	Shoes	
Jane	Books	
Jane		Max
	Books	Max

Suppose we want to delete "Max is Jane's manager". If we delete just < Jane Max > from **A** ; **B** = **A** - < Jane Max > is not closed under S : By joining < Jane Books > and < Books Max > and projecting the result on EMPLOYEE MANAGER we obtain < Jane Max > in $\pi_S(\mathbf{B}^*)$.

Thus we have to delete more than < Jane Max > from **A** to be sure that this information does not (even implicitely) exist anymore.

We encounter then two different solutions to define the result of this deletion

B

EMPLOYEE	DEPARTMENT	MANAGER
Albert	Shoes	
Paul	Shoes	
Jane	Books	

The effect of the deletion is : Max is not Jane's manager because he is not the manager of the Books department.

or

B'

EMPLOYEE	DEPARTMENT	MANAGER
Albert	Shoes	
Paul	Shoes	
	Books	Max

In this case, the meaning of the deletion is :

> Max is not Jane's manager because Jane is not employee of the Books department.

Theorem 7 :

In general, given a set of atomic facts **A** there is no unique **B** closed under S such that

(i) $\mathbf{B} \subseteq \mathbf{A}$

and (ii) no proper super set of **B** satisfies (i).

Thus, if we make a parallel with the case of the insertion and define the result of the deletion of x from **A** by the largest **B** closed under S such that $\mathbf{B} \subseteq \mathbf{A}-\{x\}$ we may have several solutions for **B**.

This ambiguity problem is also pointed out by Nicolas and Yazdanian in [NiY].
Nevertheless, there are schemata where the result of the deletion can be uniquely defined.

We say that the schema S permits to uniquely define a deletion if for any set of facts **A** closed under S and any fact x belonging to **A** there exist a unique **B** closed under S maximal such that $\mathbf{B} \subseteq \mathbf{A}-\{x\}$.

Theorem 8 :

S permits to uniquely define a deletion iff

(a) $\forall x \in S, \quad (C \in MPC(X)) \Rightarrow (|C| = 1)$

Proof :

a) Suppose that S satisfies (a) and $\mathbf{A}-\{x\} = \mathbf{A'}$ is not closed under S. Sup $\{|C| \mid \exists\ X \in S,\ C \in MPC(X)\} = 1$. Then by the theorem 5, $\pi_S(\mathbf{A'}) = \pi_S(\mathbf{A'}^*)$. Thus for each x in $\pi_S(\mathbf{A'}) - \mathbf{A'}$ there exist z in $\mathbf{A'}$ such that $z \geq x$.

Then $\mathbf{B} = \mathbf{A'} - \{y \mid \exists\ z \in \pi_S(\mathbf{A'})-\mathbf{A'} \text{ s.t. } y \geq z\}$ is closed under S.

On the other hand, any **C** closed under S and included in $\mathbf{A'}$ is disjoint from $\{y \mid \exists\ z \in \pi_S(\mathbf{A'})-\mathbf{A'} \text{ such that } y \geq z\}$.

Then **B** is the unique solution for the deletion of x from **A** and S permits to uniquely define a deletion.

b) Conversely, assume that S contains X such that

$\exists\ C \in MPC(X)\ ,\ |C| > 1$.

Let u be a fact defined on $\bigcup_{Y \in C} Y$ and **A** be :

$$\mathbf{A} = \bigcup_{\substack{Z \in S \\ Z \subseteq \bigcup_{Y \in C} Y}} \{u_{|Z}\} \qquad \mathbf{A} \text{ is closed under } S$$

We will show that there are two solutions for the result of the deletion of $x = u_{|X}$ from **A**.

C is minimal then there exist a numbering $Y_1 \dots Y_k$ of its elements such that :

$$Y_i \not\subseteq \bigcup_{j<i} Y_j \quad \forall\ i>1 \text{ and } Y_k \cap (X-(\bigcup_{j<k} (Y_j \cap X))) \neq \emptyset$$

(i.e. there is at most one element of X that is only in Y_k).

Consider the set I :

$$I = \{\mathbf{B} \mid \mathbf{B} \text{ closed under } S,\ \bigcup_{j<k} \{u_{|Y_j}\} \subseteq \mathbf{B},\ \mathbf{B} \subseteq \mathbf{A}-\{x,x'\}\}$$

where $x' = u_{|Y_k}$. I is not empty, by the choice of Y_k. I contains at least a maximal element **C** for the set inclusion.

C is also a maximal element of :

$$J = \{\mathbf{B} \mid \mathbf{B} \text{ closed under S and } \mathbf{B} \subseteq \mathbf{A}-\{x\}\}$$

If it was not the case, there would exist. **B** in **J** such that $\mathbf{C} \subset \mathbf{B}$. $\mathbf{B} \notin I$ then $x' \in \mathbf{B}$. By the definition of I $\bigcup_{Y \in C} \{u_{|Y}\} \subseteq \mathbf{B}$. As **B** is closed under S, $x \in \mathbf{B}$ which is impossible.

J has another maximal element because

$$\mathbf{B} = \bigcup_{\substack{Z \in S \\ Z \subseteq Y_k}} \{u_{|Z}\} \text{ belongs to J and } \mathbf{B} \not\subseteq \mathbf{C}.$$

Thus S does not permit to uniquely define a deletion. **Q.E.D.**

Example 9 :

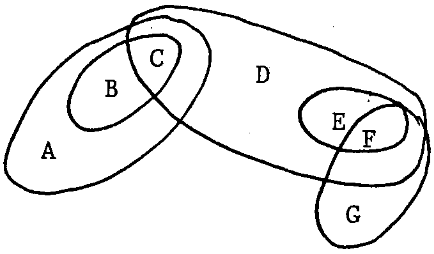
(i) Consider the schema S and the compatibility relation @ of example 3

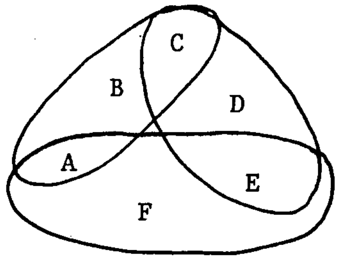
S = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR}

MPC(ClCoSR) = {{ClCoS,ClCoTR}} thus S does not permit to uniquely define a deletion.

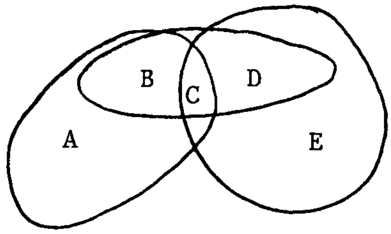
(ii) If @ satisfies X @ Y <=> $X \cap Y \neq \emptyset$ then S can be represented by a hypergraph

S permits to uniquely define a deletion :





S does not permit to uniquely define a deletion :



We can now address the problem of computing the result of a deletion when it is uniquely defined :

Theorem 9 :

When S permits to uniquely define a deletion, given **A** closed under S and x in **A**, the result of the deletion of x from **A** is :

$$B = \mathbf{A} - \{y \mid y \in \mathbf{A} \text{ and } y \geq x\}$$

Proof :

The proof follows directly from the first point of the proof of the previous theorem.

We should note that when S converges in 0 step S permits to uniquely define a deletion. Furthermore when S permits to uniquely define a deletion, S converges in one step. Thus it allows easy insertions. However there are cases where S converges in one step and S does not permit to uniquely define a deletion.

Example 10 :

Let us take the schema S and the compatibility relation @ of the example 3 :

- As we have seen in the example 9, S does not permit to uniquely define a deletion

S converges in 1 step :

S = {ClS,CoS,ClCoS,ClCoSR,ClT,CoT,ClCoT,ClCoTR}
MPC(ClCoSR) = {{ClCoS,ClCoTR}}
MPC(ClS) = {{ClCoS},{ClCoSR}}
MPC(ClCoS) = {ClCoSR}
MPC(ClT) = {{ClCoT},{ClCoTR}}
MPC(CoT) = {{ClCoT},{ClCoTR}}
MPC(ClCoTR) = {{ClCoT,ClCoSR}}

(I) is not satisfied for X = ClCoSR : (we must take only precycles having more than one element)

if C = {ClCoS}
C_Y = {ClCoSR}

There is no C' $\in$ MPC(ClCoSR) such that

$$C' \cap C_Y \neq \emptyset \quad \text{and} \quad C' \subseteq C \cup C_Y$$

Because C_Y = {ClCoSR}

(I) is not satisfied for X = ClCoTR for the same reasons.

Thus S converges in one step.

In this section we have studied the problem of defining the result of a deletion of an atomic fact from a set of facts closed under a schema. We have shown that, in the general case the result of the deletion is not uniquely defined and we have characterized the schemas where the result is unique. In these case the computation of the result is simple.

These results can be applied to justify the definition of the updates-operations for the V-relations of the Verso-machine [Ban].

An alternative approach : representing facts by a universal relation

Given a set of facts **A** closed under S, the information contained in **A** consists of all the possible facts that can be deduced from **A** i.e. $\mathbf{A}^*$. It can be represented by $\bar{\mathbf{A}}$, the minimal set of facts "information equivalent" to $\mathbf{A}^*$. By minimality we mean in terms of facts and for equivalence we follow Zaniolo's definition [Za2].

In general it is not possible to represent $\bar{\mathbf{A}}$ by a simple relation unless we use nulls : a course might have students and no teacher and vice-versa. The solutions proposed for the formal treatment of incomplete relations in database operations ([Bis], [Cod], [lip], [ImL], [Sag], [Vas], [KKU], [Za1], [Za2],...) show that the interpretation of null values adds computational complexity and semantic problems. Thus we propose a simple model, following Zaniolo in his representation of nulls :

Definition :

With a set of atomic facts **A**, we associate a relation over U as follows :

1) Extend each domain by adding a special null value :

$$\bar{D}(A) = D(A) \cup \{-\}$$

2) For each fact x in $\bar{\mathbf{A}}$, extend x to U as follows :

for $A \in U$ if $x(A)$ is not defined then $x(A)$='-'.
We call this set of tuples Rel(**A**)

Example 11 :

Let **A** be as in the example 5

A	EMPLOYEE	DEPARTMENT	MANAGER
	Albert	Shoes	
	Paul	Shoes	
	Jane	Books	
	Jane		Max
		Books	Max

then Rel(**A**) is :

EMPLOYEE	DEPARTMENT	MANAGER
Albert	Shoes	-
Paul	Shoes	-
Jane	Books	Max

We can compare this approach to that of Biskup and Bruggeman [BiB] :

From a schema $S = \{R_1 \ldots R_n\}$ and a database $d = \{r_1 \ldots r_n\}$, where each r_i is a relation on R_i, they want to define a universal relation view urv(d). urv(d) is defined on $\bigcup_{i \leq n} R_i$ and must have the same answer than d on all the queries. This condition must be valid for all intended users whether they know or not the schema.

Let us take S as a schema in our model and the compatibility relation @ defined by :

$$\forall X, Y \in \underline{O} \quad X \mathbin{@} Y \Leftrightarrow X \cap Y \neq \emptyset$$

The set of all the join path (JP in BiB) is :

$$JP = \{C \in S \ / \ C \ @ \ \text{connected}\}$$

for any subset X of $\bigcup_{i \leq n} R_i$, the set of all the join path covering X is

$$JP(X) = \{C \in JP \mid X \subseteq \bigcup_{Y \in C}\}$$

and for any E in JP

$$UE = \bigcup_{X \in E} X$$

We identify d and the set of atomic facts $\mathbf{A} = \bigcup_{i \leq n} r_i$. Then if $\mathbf{A}$ is closed under S, Rel($\mathbf{A}$) corresponds to the fictitions universal instance urv(d) when the conditions 1, 2, 3 of [BiB] are satisfied :

1 : For the consistent understanding :

$$\pi_X(urv(d)) = \bigcup_{E \in jp(X)} \pi_X(\mathop{*}_{R_i \in E} r_i)$$

jp(X) is the set of the join paths where the sophisticated user evaluate any query on X.
(We have $jp(X) \subseteq JP(X)$).

2 : urv(d) is the minimal element of :

$$\{u \mid \forall X \in \bigcup_{i \leq n} R_i,\ \pi_X(u) = \bigcup_{E \in jp(X)} \pi_X(\mathop{*}_{R_i \in E} r_i)$$

3 : For the visibility of the partial join :

$$\forall E \in JP,\ \mathop{*}_{R_i \in E} r_i \subseteq \pi_{UE}\ (urv(d))$$

When 1, 2 and 3 are satisfied then jp(X) = JP(X) for all the subsets X of $\bigcup_{i \leq n} R_i$.

Furthermore, if S converges in zero step the properties 4 and 5 of [BiB] are satisfied :

4 : For the unambiguous visibility of partial joins :

$$\forall d,\ \forall E \in JP,\ \mathop{*}_{R_i \in E} r_i = \pi_{UE}\ (urv(d))$$

5 : For an unambiguous database schema :

$$\forall E \in JP,\ \forall R_i \in S,\ \text{if } R_i \subseteq \bigcup_{X \in E} X \text{ then } R_i \in E$$

This condition is equivalent to the characterisation of "S converges in zero step".

Thus the model of Biskup and Bruggeman can be considered as a special case of our model.

Given a relation R with nulls and a schema S, we define an associated set of atomic facts $Fact_S(R)$ as follows :

$$Fact_S(R) = \{x \mid (\exists\ X \in S) \wedge (\exists\ y \in R,\ y(A) \neq \text{'-'} \quad \forall A \in X) \wedge (x = y|_X)\}$$

$Fact_S(R)$ is the total projection of R over the schema S.

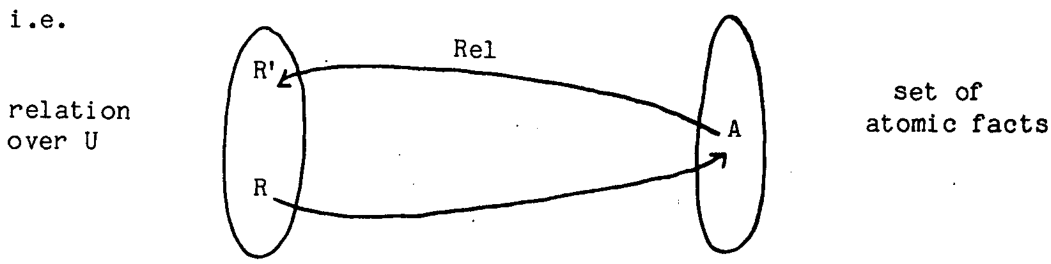
We do not want to loose or to add information while transforming a relation into a set of atomic facts or vice-versa. This leads us to define "decomposability" for relations :

Definition :

Let R be a relation eventually with null values. We say that R has a lossless decomposition (LLD) with respect to S iff

$$R = Rel(Fact_S(R))$$

i.e.



R has LLD w.r.t.S iff R=R'

Note that :

If S is @ connected then for any relation R without '-' we have $R \subseteq Rel(Fact_S(R))$.

The two notions : "closed under S" for a set of atomic facts and "LLD w.r.t. S" for a relation are related as follows :

Theorem 10 :

If R is "LLD w.r.t. S then $Fact_S(R)$ is closed under S.

Proof :

a) Let us consider the composition of operations :

$$Fact_S \circ Rel : \{\text{set of atomic facts}\} \rightarrow \{\text{set of atomic facts}\}$$

To apply Rel to a set of atomic facts

1) compute $\bar{\mathbf{A}}$

2) complete eventually the elements of $\mathbf{A}$ by '-' to obtain an incomplete relation over U

To apply $Fact_S$ to an incomplete relation we :

3) eliminate the '-' in the tuples

4) project upon S the set of facts resulting of (3).

Thus we can eliminate (2) and (3) in the process. We have : for any set of atomic facts

$$Fact_S(Rel(\mathbf{A})) = \pi_S(\bar{\mathbf{A}})$$

b) Suppose R has LLD wit respect to S

then $Fact_S(R) = Fact_S(Rel(Fact_S(R)))$

thus $Fact_S(R) = \pi_S((\overline{Fact_S(R))}$ by (a) Q.E.D.

Then if R is LLD w.r.t. S, R and $Fact_S(R)$ represent the same information.

But by knowing that $Fact_S(R)$ is closed under S we cannot ensure that R is LLD w.r.t. S.

Nevertheless we have :

Theorem 11 :

If **A** is closed under S then Rel(**A**) is LLD w.r.t. S

Proof :

Trivial by using (a) of the previous proof.

Thus if **A** is closed under S **A** can be represented by Rel(**A**) without any modification of the information.

The notion of "lossless decomposition with respect to S" is stronger than the notion of join dependency :

Theorem 12 :

If R is total (i.e. without nulls) and has LLD w.r.t. s then for any subset C of S which is @ connected and covers U, R satisfies the join dependency *C.

Proof :

Consider a total relation R which satisfies $R=Rel(Fact_S(R))$ and an @ connected subset C of S covering U.

C is @ connected, thus there exist a numbering $X_1 \ldots X_n$ of the elements of C such that

$$X_i \; @ \; (X_1 \cup \ldots \cup X_{i-1}) \qquad \forall \, i \in \{2 \ldots n\}$$

Let x be in $*_{X \in C} R[X]$ (R[X] is the projection of R on X)

$$x \in R[X_1] * \ldots * R[X_n]$$

Therefore there exist $x_1 \ldots x_n$ each x_i defined on X_i such that $x = (\ldots(x_1 * x_2)\ldots * x_n)$ for our join.
x is defined on U because C covers U

then $x \in Rel(Fact_S(R)) = R$ Q.E.D.

However the converse is not true as the following example shows

Example 12 :

Let U be {A,B,C,D,E}
S = {AB,BC,BDE,CDE}

and @ defined as follows :

@ satisfies

- if $X \subseteq Y$ then X @ Y
- if X @ Y then Y @ X
- AB @ BC
- BC @ BDE
- ABC @ CDE

There are two $C \subseteq S$, @ connected covering U

$C_1 = \{AB, BC, CDE\}$

and $C_2 = \{AB, BC, CDE, BDE\}$

R =

A	B	C	D	E
0	0	1	0	1
0	0	0	0	0

satisfies $R = \underset{C_1}{*} R = \underset{C_2}{*} R$

But :

$Rel(Fact_S(R))$ is :

A	B	C	D	E
0	0	1	0	1
0	0	0	0	0
-	0	1	0	0
-	0	0	0	1

Thus R has not LLD w.r.s. S

The constraint introduced "having LLD w.r.t. S" is not equivalent in general to a set of join dependencies : it may appear dandling tuples in $Rel(Fact_S(R))$.

CONCLUSION

We have described a model that represents facts over a set of attributes.

This model, "set of facts closed under a schema" allows two deduction rules :

- deduction by join. The compatibility relation determines the possible joins.

- deduction by projection. The only possible projections are the projections on the elements of the schema.

In this framework, we have studied update operations.
This yields two problems :

- the definition of the result of an update
- the effective computation of this result.

This led us to the characterisation of schemata where the two problems are solved

- in the case of an insertion

and

- in the case of a deletion

We have also interpreted this model in terms of a single relation having null values. We have then introduced a "decomposition constraint" : having LLD w.r.t. a schema, which can be compared to a set of join dependencies.

APPENDIX

Proof of the theorem :

S converges in 1 step iff
(I) is not satisfied

(I) There exist X in S, C_X in MPC(X) and $C \subset C_X$, $C \neq \emptyset$ and for each Y in C, C_Y in MP(Y) such that if there exist $C' \in MPC(X)$ such that

$$C' \subseteq C_X \cup (\bigcup_{Y \in C} C_Y)$$

and

$$C' \cap (\bigcup_{Y \in C} C_Y) \neq \emptyset$$

then

either $\exists\, Y \in C,\ Z \in (C_X - \{Y\}) \cup \{X\}$
such that $Z \cap Y \neq Z \cap (\bigcup_{V \in C' \cap C_Y} V)$

or

$|C| \geq 2$ and $\exists$ Y, Z in C, $Y \neq Z$ such that
$Z \cap Y \neq (\bigcup_{V \in C' \cap C_Y} V) \cap (\bigcup_{V \in C' \cap C_Z} V)$

(a) Suppose that S does not converge in one step. There exist a set $\mathbf{A}$ of atomic facts and an atomic fact x such that

$$x \in \pi_S(\mathbf{A}^{*})^{*}) - \pi_S(\mathbf{A}^{*}) \qquad x \text{ is defined on } X$$

Thus there exist $C = \{X_1 \ldots X_k\}$ in $MPC(X)$ $x_1 \ldots x_k$ in $\pi_S(\mathbf{A}^{*})$ each x_i defined on X_i such that

$$x \leq (\ldots(x_1 * x_2)\ldots * x_k)$$

each x_i is in $\pi_S(\mathbf{A}^{*})$ thus for each i there exist C_i in $MPC(X_i) \cup \{X_i\}$, $C_i = \{Z_{1(i)},\ldots,Z_{k_i(i)}\}$ and $z_{1(i)},\ldots z_{k_i(i)}$ in $\mathbf{A}$ such that

$$x_i \leq (\ldots(z_{1(i)} * z_{2(i)})\ldots * z_{k_i(i)})$$

$x \notin \pi_S(\mathbf{A}^{*})$ then for any sequence $y_1 \ldots y_m$ from $\bigcup_{i \leq k}(z_{1(i)},\ldots z_{k_i(i)})$

- either $(\ldots(y_1 * y_2)\ldots * y_m)$ is not defined

It happens when :

. $\{Y_1 \ldots Y_m\}$ is not @ connected

or

. $\exists$ ℓ, p $\in \{1\ldots m\}$, $\exists$ $A \in U$ such that

$y_\ell(A) \neq y_p(A)$

$(\ldots(x_1 * x_2)\ldots * x_k)$ is defined then there exist x_i and x_j such that

$A \notin X_i \cap X_j$

but $y_\ell \in \{z_{1(i)} \ldots z_{k_i(i)}\}$ and $y_p \in \{z_{1(j)} \ldots z_{k_j(j)}\}$

Thus

$Y_\ell \cap Y_p \neq X_i \cap X_j$

- or $(x \leq (\ldots(y_1 * y_2)\ldots * y_m))$ is false

It happens when :

$X \not\subset U_{i \leq m} Y_i$

or

$X \subset U_{i \leq m} Y_i$ and $\exists$ j $\in \{1\ldots m\}$, $A \in Y_j$

such that $x(A) \neq y_j(A)$

Thus if we take X, $C_X = \{X_1 \ldots X_k\}$, $C = \{X_i \mid C_i \neq \{X_i\}\}$ and for each Y in C $C_Y = C_i$ if $Y = X_i$, S satisfies (I).

(b) Conversely let us assume that S satisfies (I).

$C = \{X_1 \ldots X_k\}$.

Consider the set of atomic facts **B** defined as follows :

For each X_i in C and each V in C_{X_i},

$x_{i,v} : V \rightarrow U_{A \in V} D(A)$

$X_{i,V} \in$ **B**

$A \in V \cap X_i \rightarrow 0$

$A \in V - X_i \rightarrow i$

For each Z in C_X-C,

$$y_Z : Z \to \underset{A \to 0}{U_{A \in Z}} D(A) \qquad y_Z \in \mathbf{B}$$

And that is all.

(We have supposed for the sake of simplicity that $\mathbf{N}$ is included in D(A) for any attribute $\mathbf{A}$).

By joining all the $x_{i,V}$ for V in C_{X_i} and projecting the result on X_i, we obtain :

$$z_i : X_i \to \underset{A \to 0}{U_{A \in X_i}} D(A) \qquad z_i \in \pi_S(\mathbf{B}^*)$$

By joining all the z_i for X_i in C and the y_Z for Z in C_X-C and projecting the result on X, we obtain :

$$x : X \to \underset{A \to 0}{U_{A \in X}} D(A) \qquad x \in \pi_S(\pi_S(\mathbf{B}^*)^*)$$

x cannot belong to $\pi_S(\mathbf{B}^*)$:

If it was the case there would exist a sequence $u_1 \ldots u_m$ f elements of $\mathbf{B}$ such that

$$x \leq (\ldots(u_1 * u_2)\ldots * u_m) .$$

Then there would exist C' in MPC(X) such that

$$C' \subseteq (C_X - C) \cup (\bigcup_{Y \in C} C_Y) \text{ and } C' \cap (\bigcup_{Y \in C} C_Y) \neq \emptyset$$

by the choice of $\mathbf{B}$.

S satisfies (I) then

either $\exists\, Y \in C,\ Z \in (C_X - \{Y\}) \cup \{X\}$ s.t.

$Z \cap Y \neq Z \cap \left(\bigcup_{V \in C' \cap C_Y} V \right)$

thus

. either $\exists$ z, v in $\{u_1 \ldots u_m\}$ s.t. z is defined on Z and v is defined on $V \in C_Y \cap C$ s.t $V \cap Z \neq Y \cap Z$ and then u_i and u_j are not compatible

or $\exists$ v in $\{u_1 \ldots u_m\}$ defined on $V \in C_Y\ \ C$ s.t $V \cap X \neq Y \cap X$ and then $v(A) \neq 0$ for an A in X.

or $\exists\, Y, Z \in C$ s.t $Z \cap Y \neq \left(\bigcup_{V \in C' \cap C_Y} V \right) \left(\bigcup_{V \in C' \cap C_Z} V \right)$ then $\exists$ v, w in $\{u_1 \ldots u_m\}$ not compatible :

v is defined on $V \in C \cap C_Y$

s.t $V\ \ W \neq Y \cap Z$

and w is defined on $W \in C \cap C_Z$.

Thus $\mathbf{B}^1 \neq \mathbf{B}^S$ and S does not converge in one step. Q.E.D.

REFERENCES

[Ban] F. Bancilhon & al. : "VERSO : A relational Back End Data Base Machine", International Workshop on Database Machines, San Diego, 1982.

[Ber] C. Berge : "Graphe et hypergraphes", Dunod, Paris, 1973.

[BiB] J. Biskup, H.H. Bruggemann : "Universal Relation Views : A Pragmatic Approach", International Conference on Very large Data Base, Florence, 1983.

[Bis] J. Biskup : "A Formal Approach to Null Values in Database Relations", Formal Bases for Data Bases, Toulouse, Décembre 1979.

[Cod] E.F. Codd : "Extending the Database Relational Model to Capture More Meaning", ACM Transaction on Database System, 4, 1979.

[ImL] T. Imielinski, W.Jr. Lipski : "On Representing Incomplete Information in a Relational Database", International Conference on Very Large Data Base, Cannes, 1981.

[KKU] H.F. Korth, G. Kuper, J.D. Ullman : "System/U : A Database System based on the Universal Relation Assumption", Research Report No STAN-CS-82-944, Stanford University, 1983.

[Lip] W.Jr. Lipski : "On Database with Incomplete Information", JACM 28, 1981.

[MRW] D. Maier, D. Rozenshtein, D. Warren : "Window on the world", ACM SIGMOD International Conference on Management of Data, San Jose, 1983.

[NiY] J.M. Nicolas, K. Yazdanian : "An Outline of BDGEN : A Deductive DBMS", IFIP, Paris, 1983.

[Sag] Y. Sagiv : "Can we Use the Universal Instance Assumption without using Nulls ?", ACM SIGMOD International Conference on Management of Data, 1981.

[Sci] E. Sciore : "The Universal Instance and Database Design", Technnical Report 271, Princeton University, 1980.

[Vas] Y. Vassiliou : "A Formal Treatment of Imperfect Information in Database Management", Technical Report CRSG-123, Toronto, 1980.

[Ve1] A. Verroust : "Characterisation of Well-behaved Database Schemata and their Update Semantics", International Conference of Very Large Data Base, Florence, 1983.

[Ve2] A. Verroust : "Schémas bien formés et opérations de mise à jour", Thèse de 3e cycle, LRI, Université Paris Sud, Orsay, 1983.

[Za1] C. Zaniolo : "Relational Views in a Database System Support for Queries", IEEE, Computer Society Computer Software and Applications Conference, Chicago, 1977.

[Za2] C. Zaniolo : "Database Relation with Null Values", extended abstract, ACM SIGACT SIGMOD Conference on Principles of Database Systems, 1982.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique